SECOND EDITION

DATA CENTER HANDBOOK

PLAN, DESIGN, BUILD, AND OPERATIONS OF A SMART DATA CENTER

HWAIYU GENG P.E

WILEY

DATA CENTER HANDBOOK

DATA CENTER HANDBOOK

Plan, Design, Build, and Operations of a Smart Data Center

Second Edition

HWAIYU GENG, P.E.
Amica Research
Palo Alto, California, United States of America

WILEY

This second edition first published 2021

Edition History
John Wiley & Sons, Inc. (1e, 2015)

Registered Office
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

Editorial Office
111 River Street, Hoboken, NJ 07030, USA

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data

Names: Geng, Hwaiyu, editor.
Title: Data center handbook : plan, design, build, and operations of a
smart data center / edited by Hwaiyu Geng.
Description: 2nd edition. | Hoboken, NJ : Wiley, 2020. | Includes index.
Identifiers: LCCN 2020028785 (print) | LCCN 2020028786 (ebook) | ISBN
9781119597506 (hardback) | ISBN 9781119597544 (adobe pdf) | ISBN
9781119597551 (epub)
Subjects: LCSH: Electronic data processing departments–Design and
construction–Handbooks, manuals, etc. | Electronic data processing
departments–Security measures–Handbooks, manuals, etc.
Classification: LCC TH4311 .D368 2020 (print) | LCC TH4311 (ebook) | DDC
004.068/4–dc23
LC record available at https://lccn.loc.gov/2020028785
LC ebook record available at https://lccn.loc.gov/2020028786

Cover Design: Wiley
Cover Image: Particle earth with technology network over Chicago Cityscape © Photographer is my life. / Getty Images, front cover
icons © Macrovector / Shutterstock except farming icon © bioraven / Shutterstock

Set in 10/12pt Times by SPi Global, Pondicherry, India

10 9 8 7 6 5 4 3 2 1

To "Our Mothers Who Cradle the World" and To "Our Earth Who Gives Us Life."

BRIEF CONTENTS

ABOUT THE EDITOR/AUTHOR

Hwaiyu Geng, CMfgE, P.E., is a principal at Amica Research (Palo Alto, California, USA) promoting green technological and manufacturing programs. He has had over 40 years of diversified technological and management experience, worked with Westinghouse, Applied Materials, Hewlett-Packard, Intel, and Juniper Network on international hightech projects. He is a frequent speaker in international conferences, universities, and has presented many technical papers. A patent holder, Mr. Geng is also the editor/author of the *Data Center Handbook (2ed), Manufacturing Engineering Handbook (2ed), Semiconductor Manufacturing Handbook (2ed), and the IoT and Data Analytics Handbook*.

TECHNICAL ADVISORY BOARD

LIST OF CONTRIBUTORS

Ken Baudry, K.J. Baudry, Inc., Atlanta, Georgia, United States of America

Sergio Bermudez, IBM TJ Watson Research Center, Yorktown Heights, New York, United States of America

David Bonneville, Degenkolb Engineers, San Francisco, California, United States of America

David Cameron, Operational Intelligence Ltd, London, United Kingdom

Ronghui Cao, College of Information Science and Engineering, Hunan University, Changsha, China

Nicholas H. Des Champs, Munters Corporation, Buena Vista, Virginia, United States of America

Christopher Chen, Jensen Hughes, College Park, Maryland, United States of America

Chris Crosby, Compass Datacenters, Dallas, Texas, United States of America

Chris Curtis, Compass Datacenters, Dallas, Texas, United States of America

Sean S. Donohue, Jensen Hughes, Colorado Springs, Colorado, United States of America

Keith Dunnavant, Munters Corporation, Buena Vista, Virginia, United States of America

Mark Fisher, Munters Corporation, Buena Vista, Virginia, United States of America

Sophia Flucker, Operational Intelligence Ltd, London, United Kingdom

Hubertus Franke, IBM, Yorktown Heights, New York, United States of America

Ajay Garg, Intel Corporation, Hillsboro, Oregon, United States of America

Chang-Hsin Geng, Supermicro Computer, Inc., San Jose, California, United States of America

Hwaiyu Geng, Amica Research, Palo Alto, California, United States of America

Hendrik Hamann, IBM TJ Watson Research Center, Yorktown Heights, New York, United States of America

Sarah Hanna, Facebook, Fremont, California, United States of America

Skyler Holloway, Facebook, Menlo Park, California, United States of America

Ching-I Hsu, Raritan, Inc., Somerset, New Jersey, United States of America

Dongmei Huang, Beijing Rainspur Technology, Beijing, China

Robert Hunter, AlphaGuardian, San Ramon, California, United States of America

Phil Isaak, Isaak Technologies Inc., Minneapolis, Minnesota, United States of America

Alexander Jew, J&M Consultants, Inc., San Francisco, California, United States of America

Masatoshi Kajimoto, ISACA, Tokyo, Japan

Levente Klein, IBM TJ Watson Research Center, Yorktown Heights, New York, United States of America

Bill Kosik, DNV Energy Services USA Inc., Chicago, Illinois, United States of America

Nuoa Lei, Northwestern University, Evanston, Illinois, United States of America

Bang Li, Eco Atlas (Shenzhen) Co., Ltd, Shenzhen, China

Chung-Sheng Li, PricewaterhouseCoopers, San Jose, California, United States of America

Kenli Li, College of Information Science and Engineering, Hunan University, Changsha, China

Keqin Li, Department of Computer Science, State University of New York, New Paltz, New York, United States of America

Weiwei Lin, School of Computer Science and Engineering, South China University of Technology, Guangzhou, China

Chris Loeffler, Eaton, Raleigh, North Carolina, United States of America

Fernando Marianno, IBM TJ Watson Research Center, Yorktown Heights, New York, United States of America

Eric R. Masanet, Northwestern University, Evanston, Illinois, United States of America

Robert E. Mcfarlane, Shen Milsom & Wilke LLC, New York, New York, United States of America Marist College, Poughkeepsie, New York, United States of America ASHRAE TC 9.9, Atlanta, Georgia, United States of America ASHRAE SSPC 90.4 Standard Committee, Atlanta, Georgia, United States of America

Malik Megdiche, Schneider Electric, Eybens, France

Christopher O. Muller, Muller Consulting, Lawrenceville, Georgia, United States of America

Liam Newcombe, Romonet, London, United Kingdom

Jay Park, Facebook, Fremont, California, United States of America

Robert Pekelnicky, Degenkolb Engineers, San Francisco, California, United States of America

Robert Reid, Panduit Corporation, Tinley Park, Illinois, United States of America

Mark Seymour, Future Facilities, London, United Kingdom

Dror Shenkar, Intel Corporation, Israel

Ed Spears, Eaton, Raleigh, North Carolina, United States of America

Richard T. Stuebi, Institute for Sustainable Energy, Boston University, Boston, Massachusetts, United States of America

Mark Suski, Jensen Hughes, Schaumburg, Illinois, United States of America

Zhuo Tang, College of Information Science and Engineering, Hunan University, Changsha, China

Robert Tozer, Operational Intelligence Ltd, London, United Kingdom

John Weale, The Integral Group, Oakland, California, United States of America

Joseph Weiss, Applied Control Solutions, Cupertino, California, United States of America

Beth Whitehead, Operational Intelligence Ltd, London, United Kingdom

Jan Wiersma, EVO Venture Partners, Seattle, Washington, United States of America

Wentai Wu, Department of Computer Science, University of Warwick, Coventry, United Kingdom

Chao Yang, Chongqing University, Chongqing, China

Ligong Zhou, Raritan, Inc., Beijing, China

FOREWORD (1)

The digitalization of our economy requires data centers to continue to innovate to meet the new needs for connectivity, growth, security, innovation, and respect for the environment demanded by organizations. Every phase of life is putting increased pressure on data centers to innovate at a rapid pace. Explosive growth of data driven by 5G, Internet of Things (IoT), and Artificial Intelligence (AI) is changing the way data is stored, managed, and transferred. As this volume grows, data and applications are pulled together, requiring more and more computing and storage resources. The question facing data center designers and operators is how to plan for the future that accomplishes the *security, flexibility, scalability, adaptability, and sustainability needed* to support business requirements.

With this explosion of data, companies need to think more carefully and strategically about how and where their data is stored, and the security risks involved in moving data. The sheer volume of data creates additional challenges in protecting it from intrusions. This is probably one of the most important concerns of the industry – how to protect data from being hacked and being compromised in a way that would be extremely damaging to their core business and the trust of their clients.

Traditional data centers must deliver a degree of *scalability* to accommodate usage needs. With newer technologies and applications coming out daily, it is important to be able to morph the data center into the needs of the business. It is equally important to be able to integrate these technologies in a timely manner that does not compromise the strategic plans of the business. With server racks getting denser every few years, the rest of the facility must be prepared to support an ever increasing power draw. A data center built over the next decade must be expandable to accommodate for future technologies, or risk running out of room for support infrastructure. Server rooms might have more computing power in the same area, but they will also need more power and cooling to match. Institutions are also moving to install advanced applications and workloads related to AI, which requires high-performance computing. To date, these racks represent a very small percentage of total racks, but they nevertheless can present unfamiliar power and cooling challenges that must be addressed. The increasing interest in direct liquid cooling is in response to high-performance computing demands.

5G enables a new kind of network that is designed to connect virtually everyone and everything together including machines, objects, and devices. It will require more bandwidth, faster speeds, and lower latency, and the data center infrastructure must be *flexible and adaptable* in order to accommodate these demands. With the need to bring computing power closer to the point of connectivity, the end user is driving demand for edge data centers. Analyzing the data where it is created rather than sending it across various networks and data centers helps to reduce response latency, thereby removing a bottleneck from the decision-making process. In most cases, these data centers will be, remotely managed and unstaffed data centers. Machine learning will enable real-time adjustments to be made to the infrastructure without the need for human interaction.

With data growing exponentially, data centers may be impacted by significant increases in energy usage and carbon footprint. Hyperscalers have realized this and have increasingly used more and more *sustainable* technologies. This trend will cause others to follow and adopt some of the building technologies and use of renewables for their own data centers. The growing mandate for corporations to shift to a greener energy footprint lays the groundwork for new approaches to data center power.

The rapid innovations that are occurring inside (edge computing, liquid cooling, etc.) and outside (5G, IoT, etc.) of data centers will require careful and thoughtful analysis to design and operate a data center for the future that will serve the strategic imperatives of the business it supports. To help address the complex environment with competing forces, this second edition of the Data Center Handbook has assembled by leaders in the industry and academia to share their latest thinking on these issues. This handbook is the most comprehensive guide available to data center practitioners as well as academia.

Roger R. Schmidt, Ph.D.
Member, National Academy of Engineering
Traugott Distinguished Professor, Syracuse University
IBM Fellow Emeritus (Retired)

FOREWORD (2)

A key driver of innovation in modern industrial societies in the past two centuries is the application of what researchers call "general purpose technologies," which have far-ranging effects on the way the economy produces value. Some important examples include the steam engine, the telegraph, the electric power grid, the internal combustion engine, and most recently, computers and related information and communications technologies (ICTs).

ICTs represent the most powerful general-purpose technologies humanity has ever created. The pace of innovation across virtually all industries is accelerating, which is a direct result of the application of ICTs to increase efficiency, enhance organizational effectiveness, and reduce costs of manufacturing products. Services provided by data centers enable virtually all ICTs to function better.

This volume presents a comprehensive look at the current state of the data center industry. It is an essential resource for those working in the industry, and for those who want to understand where it is headed.

The importance of the data center industry has led to many misconceptions, the most common of which involves inflated estimates of how much electricity data centers use. The latest credible estimates for global electricity use of data centers are for 2018, from our article in *Science Magazine* in February 2020 (Masanet et al. 2020).

According to this analysis, data centers used about 0.9% of the world's electricity consumption in 2018 (down from 1.1% in 2010). Electricity use grew only 6% even as the number of compute instances, data transfers, and total data storage capacity grew to be 6.5 times, 11 times, and 26 times as large in 2018 as each was in 2010, respectively.

The industry was able to keep data center electricity use almost flat in absolute terms from 2010 to 2018 because of the adoption of best practices outlined in more detail in this volume. The most consequential of these best practices was the rapid adoption of hyperscale data centers, known colloquially as cloud computing. Computing output and data transfers increased rapidly, but efficiency also increased rapidly, almost completely offsetting growth in demand for computing services.

For those new to the world of data centers and information technology, this lesson is surprising. Even though data centers are increasingly important to the global economy, they don't use a lot of electricity in total, because innovation has rapidly increased their efficiency over time. If the industry aggressively adopts the advanced technologies and practices described in this volume, they needn't use a lot of electricity in the future, either.

I hope analysts and practitioners around the world find this volume useful. I surely will!

Jonathan Koomey, Ph.D.,
President, Koomey Analytics
Bay Area, California

FOREWORD (3)

The data center industry changes faster than any publication can keep up with. So why the "Data Center Handbook"? There are many reasons, but three stand out. First, fundamentals have not changed. Computing equipment may have dramatically transformed in processing power and form factor since the first mainframes appeared, but it is still housed in secure rooms, it still uses electricity, it still produces heat, it must still be cooled, it must still be protected from fire, it must still be connected to its users, and it must still be managed by humans who possess an unusual range of knowledge and an incredible ability to adapt to fast changing requirements and conditions. Second, new people are constantly entering what, to them, is this brave new world. They benefit from having grown up with a computer (i.e., "smart phone") in their hands, but are missing the contextual background behind how it came to be and what is needed to keep it working. Whether they are engineers designing their first enterprise, edge computing, hyperscale or liquid cooled facility, or IT professionals given their first facility or system management assignment within it, or are students trying to grasp the enormity of this industry, having a single reference book is far more efficient than plowing through the hundreds of articles published in multiple places every month. Third, and perhaps even more valuable in an industry that changes so rapidly, is having a volume that also directs you to the best industry resources when more or newer information is needed.

The world can no longer function without the computing industry. It's not regulated like gas and electric, but it's as critical as any utility, making it even more important for the IT industry to maintain itself reliably. When IT services fail, we are even more lost than in a power outage. We can use candles to see, and perhaps light a fireplace to stay warm. We can even make our own entertainment! But if we can't get critical news, can't pay a bill on time, or can't even make a critical phone call, the world as we now know it comes to a standstill. And that's just the personal side. Reliable, flexible, and highly adaptable computing facilities are now necessary to our very existence. Businesses have gone bankrupt after computing failures. In health care and public safety, the availability of those systems can literally spell life or death.

In this book you will find chapters on virtually every topic you could encounter in designing and operating a data center – each chapter written by a recognized expert in the field, highly experienced in the challenges, complexities, and eccentricities of data center systems and their supporting infrastructures. Each section has been brought up-to-date from the previous edition of this book as of the time of publication. But as this book was being assembled, the COVID 19 pandemic occurred, putting unprecedented demands on computing systems overnight. The industry reacted, proving beyond question its ability to respond to a crisis, adapt its operating practices to unusual conditions, and meet the inordinate demands that quickly appeared from every industry, government, and individual. A version of the famous Niels Bohr quote goes, "An expert is one who, through his own painful experience, has learned all the mistakes in a given narrow field." Adherence to the principles and practices set down by the authors of this book, in most cases gained over decades through their own personal and often painful experiences, enabled the computing industry to respond to that crisis. It will be the continued adherence to those principles, honed as the industry continues to change and mature, that will empower it to respond to the next critical situation. The industry should be grateful that the knowledge of so many experts has been assembled into one volume from which everyone in this industry can gain new knowledge.

Robert E. McFarlane
Principal, Shen Milsom & Wilke, LLC
Adjunct Faculty – Marist College, Poughkeepsie, NY

PREFACE DATA CENTER HANDBOOK (SECOND EDITION, 2021)

As Internet of Things, data analytics, artificial intelligence, 5G, and other emerging technologies revolutionize the services and products companies, the demand for computing power grows along the value chain between edge and cloud. Data centers need to improve and advance continuously to fulfill this demand.

To meet the megatrends of globalization, urbanization, demographic changes, technology advancements, and sustainability concerns, C-suite executives and technologists must work together in preparing strategic plans for deploying data centers around the world. Workforce developments and the redundancy of infrastructures required between edge and cloud need to be considered in building and positioning data centers globally.

Whether as a data center designer, user, manager, researcher, professor, or student, we all face increasing challenges in a cross-functional environment. For each data center project, we should ask, what are the goals, and work out "How to Solve It."[1] To do this, we can employ a 5W1H[2] approach applying data analytics and nurture the creativity that is needed for invention and innovation. Additionally, a good understanding of the anatomy, ecosystem, and taxonomy, of a data center will help us master and solve this complex problem.

The goal of this Data Center *Handbook* is to provide readers with the essential knowledge that is needed to plan, build, and operate a data center. This handbook embraces both emerging technologies and best practices. The handbook is divided into four parts:

Part I: *Data Center Overview and Strategic Planning* that provides an overview of data center strategic planning, while considering the impact of emerging technologies. This section also addresses energy demands, sustainability, edge to cloud computing, financial analysis, and managing data center risks.

Part II: *Data Center Technologies* that covers technologies applicable to data centers. These include software-defined applications, infrastructure, resource management, ASHRAE[3] thermal guidelines, design of energy-efficient IT equipment, wireless sensor network, telecommunication, rack level and server level cooling, data center corrosion and contamination control, cabling, cybersecurity, and data center microgrids.

Part III: *Data Center Design and Construction* that discusses plan, design, and construction of a data center that includes site selection, facility layout and rack floor plan, mechanical design, electrical design, structural design, fire protection, computational fluid dynamics, and project management for construction.

Part IV: *Data Center Operations* that covers data center benchmarking, data center infrastructure management (DCIM), energy efficiency assessment, and AI applications for data centers. This section also reviews lessons imparted from disasters, and includes mitigation strategies to ensure business continuity.

[1] Polya, G. *How to Solve It*. Princeton: Princeton University Press; 1973.

[2] The 5W1H are "Who, What, When, Where, Why, and How."

[3] ASHRAE is the American Society of Heating, Refrigerating, and Air-Conditioning Engineers.

Containing 453 figures, 101 tables and 17 pages in the index section, this second edition of *Data Center Handbook* is a single-volume, comprehensive guide to this field. The handbook covers the breadth and depth of data center technologies, and includes the latest updates from this fast-changing field. It is meant to be a relevant, practical, and enlightening resource for global data center practitioners, and will be a useful reference book for anyone whose work requires data centers.

Hwaiyu Geng, CMfgE, P.E.
Palo Alto, California, United States of America

PREFACE DATA CENTER HANDBOOK (FIRST EDITION, 2015)

Designing and operating a sustainable data center (DC) requires technical knowledge and skills from strategic planning, complex technologies, available best practices, optimum operating efficiency, disaster recovery, and more.

Engineers and managers all face challenges operating across functionalities, for example, facilities, IT, engineering, and business departments. For a mission-critical, sustainable DC project, we must consider the following:

- What are the goals?
- What are the givens?
- What are the constraints?
- What are the unknowns?
- Which are the feasible solutions?
- How is the solution validated?

How does one apply technical and business knowledge to develop an optimum solution plan that considers emerging technologies, availability, scalability, sustainability, agility, resilience, best practices, and rapid time to value? The list can go on and on. Our challenges may be as follows:

- To prepare a strategic location plan
- To design and build a mission-critical DC with energy-efficient infrastructure
- To apply best practices thus consuming less energy
- To apply IT technologies such as cloud and virtualization and
- To manage DC operations thus reducing costs and carbon footprint

A good understanding of DC components, IT technologies, and DC operations will enable one to plan, design, and implement mission-critical DC projects successfully. The goal of this handbook is to provide DC practitioners with essential knowledge needed to implement DC design and construction, apply IT technologies, and continually improve DC operations. This handbook embraces both conventional and emerging technologies, as well as best practices that are being used in the DC industry. By applying the information contained in the handbook, we can accelerate the pace of innovations to reduce energy consumption and carbon emissions and to "Save Our Earth Who Gives Us Life."

The handbook covers the following topics:

- DC strategic planning
- Hosting, colocation, site selection, and economic justifications
- Plan, design, and implement a mission-critical facility
- IT technologies including virtualization, cloud, SDN, and SDDC
- DC rack layout and MEP design
- Proven and emerging energy efficiency technologies
- DC project management and commissioning
- DC operations
- Disaster recovery and business continuity

Each chapter includes essential principles, design, and operations considerations, best practices, future trends, and further readings. The principles cover fundamentals of a technology and its applications. Design and operational considerations include system design, operations, safety, security, environment issues, maintenance, economy, and best practices. There are useful tips for planning, implementing, and controlling operational processes. The future trends and further reading sections provide visionary views and lists of relevant books, technical papers, and websites for additional reading.

This *Data Center Handbook* is specifically designed to provide technical knowledge for those who are responsible for the design, construction, and operation of DCs. It is also useful for DC decision makers who are responsible for strategic decisions regarding capacity planning and technology investments. The following professionals and managers will find this handbook to be a useful and enlightening resource:

- C-level Executives (Chief Information Officer, Chief Technology Officer, Chief Operating Officer, Chief Financial Officer)
- Data Center Managers and Directors
- Data Center Project Managers
- Data Center Consultants
- Information Technology and Infrastructure Managers
- Network Operations Center and Security Operations Center Managers
- Network, Cabling, and Communication Engineers
- Server, Storage, and Application Managers
- IT Project Managers
- IT Consultants
- Architects and MEP Consultants
- Facilities Managers and Engineers
- Real Estate Portfolio Managers
- Finance Managers

This *Data Center Handbook* is prepared by more than 50 world-class professionals from eight countries around the world. It covers the breadth and depth of DC planning, designing, construction, and operating enterprise, government, telecommunication, or R&D Data Centers. This *Data Center Handbook* is sure to be the most comprehensive single-source guide ever published in its field.

Hwaiyu Geng, CMfgE, P.E.
Palo Alto, California, United States of America

ACKNOWLEDGEMENTS
DATA CENTER HANDBOOK (SECOND EDITION, 2021)

The Data Center Handbook is a collective representation of an international community with scientists and professionals comprising 58 experts from six countries around the world.

I am very grateful to the members of the Technical Advisory Board for their diligent reviews of this handbook, confirming technical accuracy while contributing their unique perspectives. Their guidance has been invaluable to ensure that the handbook can meet the needs of a broad audience.

I gratefully acknowledge to the contributors who share their wisdom and valuable experiences in spite of their busy schedules and personal lives.

Without the trust and support from our team members, this handbook could not have been completed. Their collective effort has resulted in a work that adds tremendous value to the data center community.

Thanks must go to the following individuals for their advice, support, and contribution:

- Nicholas H. Des Champs, Munters Corporation
- Mark Gaydos, Nlyte Software
- Dongmei Huang, Rainspur Technology
- Phil Isaak, Isaak Technologies
- Jonathan Jew, J&M Consultants
- Levente Klein, IBM
- Bill Kosik, DNV Energy Services USA Inc.
- Chung-Sheng Li, PricewaterhouseCoopers
- Robert McFarlane, Shen Milsom & Wilke
- Malik Megdiche, Schneider Electric
- Christopher Muller, Muller Consulting
- Liam Newcombe, Romonet Ltd.
- Roger Schmidt, National Academy of Engineering Member
- Mark Seymour, Future Facilities
- Robert Tozer, Operational Intelligence
- John Weale, the Integral Group.

This book benefited from the following organizations and institutes and more:

- 7×24 Exchange International
- ASHRAE (American Society of Heating, Refrigerating, and Air Conditioning Engineers)
- Asetek
- BICSI (Building Industry Consulting Service International)
- Data Center Knowledge
- Data Center Dynamics
- ENERGY STAR (the U.S. Energy Protection Agency)
- European Commission Code of Conduct
- Federal Energy Management Program (the U.S. Dept. of Energy)
- Gartner
- Green Grid, The
- IDC (International Data Corporation)
- Japan Data Center Council
- LBNL (the U.S. Dept. of Energy, Lawrence Berkeley National Laboratory)
- LEED (the U.S. Green Building Council, Leadership in Energy and Environmental Design)
- McKinsey Global Institute
- Mission Critical Magazine
- NIST (the U.S. Dept. of Commerce, National Institute of Standards and Technology)

- NOAA (the U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration)
- NASA (the U.S. Dept. of Interior, National Aeronautics and Space Administration)
- Open Compute Project
- SPEC (Standard Performance Evaluation Corporation)
- TIA (Telecommunications Industry Association)
- Uptime Institute/451 Research

Thanks are also due to Brett Kurzman and staff at Wiley for their support and guidance.

My special thanks to my wife, Limei, my daughters, Amy and Julie, and my grandchildren, Abby, Katy, Alex, Diana, and David, for their support and encouragement while I was preparing this book.

Hwaiyu Geng, CMfgE, P.E.
Palo Alto, California, United States of America

ACKNOWLEDGMENTS
DATA CENTER HANDBOOK (FIRST EDITION, 2015)

The Data Center Handbook is a collective representation of an international community with scientists and professionals from eight countries around the world. Fifty-one authors, from data center industry, R&D, and academia, plus fifteen members at Technical Advisory Board have contributed to this book. Many suggestions and advice were received while I prepared and organized the book.

I gratefully acknowledge the contributors who dedicated their time in spite of their busy schedule and personal lives to share their wisdom and valuable experience.

I would also like to thank the members at Technical Advisory Board for their constructive recommendations on the structure of this handbook and thorough peer review of book chapters.

My thanks also go to Brett Kurzman, Alex Castro, Katrina Maceda at Wiley and F. Pascal Raj at SPi Global whose can do spirit and teamwork were instrumental in producing this book.

Thanks and appreciation must go to the following individuals for their advice, support, and contributions:

Sam Gelpi, Hewlett-Packard Company

Dongmei Huang, Ph.D., Rainspur Technology, China

Madhu Iyengar, Ph.D., Facebook, Inc.

Jonathan Jew, J&M Consultants

Jonathan Koomey, Ph.D., Stanford University

Tomoo Misaki, Nomura Research Institute, Ltd., Japan

Veerendra Mulay, Ph.D., Facebook, Inc.

Jay Park, P.E., Facebook, Inc.

Roger Schmidt, Ph.D., IBM Corporation

Hajime Takagi GIT Associates, Ltd., Japan

William Tschudi, P.E., Lawrence Berkeley National Laboratory

Kari Capone, John Wiley & Sons, Inc.

This book benefited from the following organizations and institutes:

7 × 24 Exchange International

American Society of Heating, Refrigerating, and Air Conditioning Engineers (ASHRAE)

Building Industry Consulting Service International (BICSI)

Datacenter Dynamics

European Commission Code of Conduct

The Green Grid

Japan Data Center Council

Open Compute Project

Silicon Valley Leadership Group

Telecommunications Industry Association (TIA)

Uptime Institute/451 Research

U.S. Department of Commerce, National Institute of Standards and Technology

U.S. Department of Energy, Lawrence Berkeley National Laboratory

U.S. Department of Energy, Oak Ridge National Laboratory

U.S. Department of Energy, Office of Energy Efficiency & Renewable Energy

U.S. Department of Homeland Security, Federal Emergency Management Administration

U.S. Environmental Protection Agency, ENERGY STAR Program

U.S. Green Building Council, Leadership in Energy & Environmental Design

My special thanks to my wife, Limei, my daughters, Amy and Julie, and grandchildren for their understanding, support, and encouragement when I was preparing this book.

TABLE OF CONTENTS

PART IV DATA CENTER OPERATIONS MANAGEMENT

PART I

DATA CENTER OVERVIEW AND STRATEGIC PLANNING

1

SUSTAINABLE DATA CENTER: STRATEGIC PLANNING, DESIGN, CONSTRUCTION, AND OPERATIONS WITH EMERGING TECHNOLOGIES

Hwaiyu Geng
Amica Research, Palo Alto, California, United States of America

1.1 INTRODUCTION

The earliest known use of the term "megatrend" was in 1980s published in the Christian Science Monitor (Boston). Oxford dictionary defines megatrend as "An important shift in the progress of a society." Internet searches reveal many megatrend reports that were published by major consulting firms including Accenture, Frost, KPMG, McKinsey Global Institute, PwC, etc. as well as organizations such as UN (United Nations)* and OECD (Organization for Economic Co-operation and Development [1]). One can quickly summarize key megatrends reported that include globalization, urbanization, demographic trend, technological breakthroughs, and climate changes.

Globalization: From Asia to Africa, multinational corporations are expanding their manufacturing and R&D at a faster pace and on a larger scale than ever before. Globalization widely spreads knowledge, technologies, and modern business practices at a faster space that facilitate international cooperation. Goods and services inputs are increasingly made of countries from emerging economies who join key global players. Global value chains focus on national innovation capacities and enhance national industrial specialization. Standardization, compatibility, and harmonization are even more important in a global interlaced environment.

Urbanization: Today, more than half of the world's population live in urban areas, and more people are moving to the urban areas every day. The impacts from urbanization are enormous. Demands for infrastructure, jobs, and services must be met. Problems of human health, crime, and pollution of the environment must be solved.

Demographic trend: Longer life expectancy and lower fertility rate are leading to rapidly aging populations. We must deal with increasing population, food and water shortages, and preserving natural resources. At the same time, sex discrimination, race and wealth inequalities in every part of the world must be dealt with.

Technological changes: New technologies create both challenges and opportunities. Technological breakthroughs include Internet of Things (IoT), cyber–physical systems (CPS), data analytics, artificial intelligence (AI), robotics, autonomous vehicles (AVs) (robots, drones), cloud and edge computing, and many other emerging technologies that fuel more innovative applications. These technologies fundamentally change our lifestyle and its ecosystem. Industries may be disrupted, but more inventions and innovations are nurturing.

Climate change and sustainability: Unusual patterns of droughts, floods, and hurricanes are already happening. The world is experiencing the impacts of climate change, from melting glaciers to rising sea level to extreme weather patterns. In the April 17, 2020, *Science* magazine issue, researchers examine tree rings and report that the drought from 2000 to 2018 in the southwestern of North America is among the worst "megadroughts" that have stricken the region in the last 1,200 years. The United Nation's IPCC (Intergovernmental Panel on Climate Change) reports have described increasing dangers of climate change. At the current rising rate of

* https://www.un.org/development/desa/publications/wp-content/uploads/sites/10/2020/09/20-124-UNEN-75Report-2-1.pdf

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.

greenhouse gas emissions, the global average temperature will rise by more than 3°C in the twenty-first century. Rising temperatures must be kept below 2°C before year 2050 or potential irreversible environmental changes will occur. It is imperative to find sustainable solutions and delay climate change.

This chapter will start with megatrends and emerging technologies that provide insightful roadmap of future data centers and essential elements to be included when designing and implementing a data center project.

1.1.1 Data Center Definition

Data centers are being used to orchestrate every aspect of our life that covers food, clothing, shelter, transportation, healthcare, social activities, etc. The U.S. Environmental Protection Agency defines a data center as:

- "Primarily electronic equipment used for data processing (servers), data storage (storage equipment), and communications (network equipment). Collectively, this equipment processes, stores, and transmits digital information."
- "Specialized power conversion and backup equipment to maintain reliable, high-quality power, as well as environmental control equipment to maintain the proper temperature and humidity for the ICT (information and communication technologies) equipment."

A data center could also be called data hall, data farm, data warehouse, AI lab, R&D software lab, high-performance computing lab, hosting facility, colocation, computer room, server room, etc.

An exascale data center has computing systems that perform calculation over a petaflop (a million trillion floating-point) operations. Exascale data centers are elastically configured and deployed that can meet specific workloads and be optimized for future developments in power and cooling technology.[1]

The size of a data center could range from a small closet to a hyperscale data center. The term hyperscale refers to a resilient and robust computer architecture that has the ability to increase computing ability in memory, networking, and storage resources.

Regardless of size and what it is called, all data centers perform one thing, that is, to process and deliver information.

[1] http://www.hp.com/hpinfo/newsroom/press_kits/2008/cloudresearch/fs_exascaledatacenter.pdf

1.1.2 Data Center Energy Consumption Trends

The energy consumption trend depends on a combination of factors including data traffic, emerging technologies, ICT equipment, and energy demand by infrastructure in data centers. The trend is a complicated and dynamic model. According to "United States Data Center Energy Usage Report, Lawrence Berkeley National Laboratory" (2016) by Arman Shehabi, Jonathan Koomey, et al. [2], U.S. data center electricity used by servers, storage, network equipment, and infrastructure in 2014 consumed an estimated of 70 billion kWh. That represents about 1.8% of total U.S. electricity consumption. The U.S. electricity used by data centers in 2016 was 2% of global electricity. For 70 billion kWh, it is equivalent to 8 nuclear reactors with 1,000 MW baseload each. 70 billion kWh provides enough energy for use by 5.9 million homes in 1 year.[2] It is equivalent to 50 million ton of carbon dioxide emission to the atmosphere. It is expected that electricity consumption will continue to increase and data centers must be valiantly controlled to conserve energy use.

1.2 ADVANCED TECHNOLOGIES

The United Nations predicts that the world's population of 7.8 billion people in 2020 will reach 8.5 billion in 2030 and 9.7 billion in 2050.[3] Over 50% of the world's population are Internet users that demand more uses of data centers. This section will discuss some of the important emerging technologies illustrated by its anatomy, ecosystem, and taxonomy. Anatomy defines components of a technology. Ecosystem describes who uses the technology. Taxonomy is to classify the components of a technology and their providers in different groups. With a good understanding of what is anatomy, ecosystem, and taxonomy of a technology, one can effectively apply and master the technology.

1.2.1 Internet of Things

The first industrial revolution (IR) started with the invention of mechanical powers. The second IR happened with the invention of assembly line and electrical power. The third IR came about with computers and automation. The fourth IR took place around 2014 as a result of the invention of IoT. IDC (International Data Corporation) forecasts an expected IoT market size of $1.1 trillion in 2023. By 2025, there will be 41.6 billion IoT connected devices that will generate 79.4 zettabytes (ZB) of data.

[2] https://eta.lbl.gov/publications/united-states-data-center-energy

[3] https://population.un.org/wpp/Graphs/1_Demographic%20Profiles/World.pdf

The IoT is a series of hardware coupling with software and protocols to collect, analyze, and distribute information. Using the human body as an analogy, humans have five basic senses or sensors that collect information. Nervous system acts as a network that distributes information. And the brain is accountable for storing, analyzing, and giving direction through the nervous system to five senses to execute decision. The IoT works similar to the combination of five senses, the nervous system and the brain.

1.2.1.1 Anatomy

Anatomy of the IoT comprises of all components in the following formula:

$$\begin{aligned}\text{Internet of Things} = {} & \text{Things}\left(\text{sensors/cameras/actuators}\right)\\ & + \text{edge / fog computing and AI}\\ & + \text{Wi-Fi / gateway / 5G / Internet}\\ & + \text{cloud computing / data analytics / AI}\\ & + \text{insight presentations / actions}\end{aligned}$$

Each "Thing" has a unique IPv4 or IPv6 address. A "Thing" could be a person, an animal, an AV, or alike that is interconnected at many other "Things." With increasing miniaturization and built-in AI logics, sensors are performing more computing at "edge" as well as other components in the IoT's value chain before arriving at data centers for "cloud computing." AI is embedded in every component and becomes an integral part of the IoT. This handbook considers Artificial Intelligence of Things (AIoT) the same as the IoT.

1.2.1.2 Ecosystem

There are consumer-, government-, and enterprise-facing customers within an IoT's ecosystem (Fig. 1.1). Each IoT platform contains applications that are protected by a cybersecurity system. Consumer-facing customers be composed of smart home, smart entertainment, smart health, etc. Government-facing customers are composed of smart cities, smart transportation, smart grid, etc. Enterprise-facing customers include smart retail, smart manufacturing, smart finance, etc.

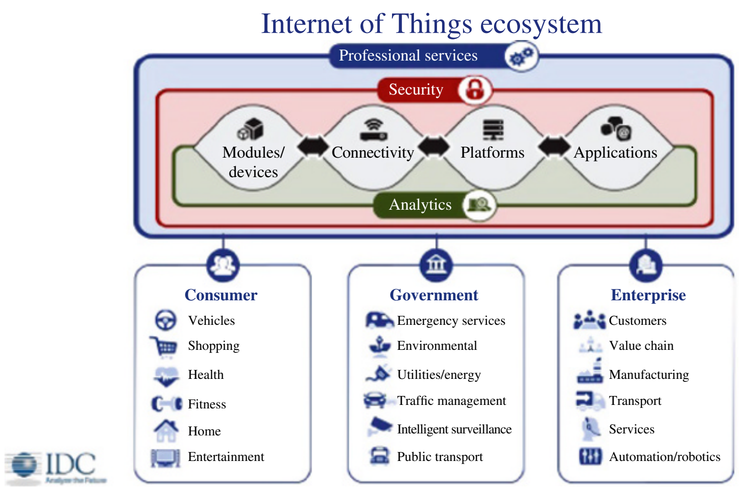
1.2.1.3 Taxonomy

Using taxonomy in a hospital as an analogy, a hospital has an admission office, medical record office, internal medicine, cardiology, neurology, radiology, medical laboratory, therapeutic services, pharmacy, nursing, dietary, etc. IoT's taxonomy encompasses suppliers who provide products, equipment, or services that cover sensors (microprocessor unit, system on chip, etc.), 5G, servers, storage, network, security, data analytics, AI services, industry solutions, etc.

The Industrial IoT (IIoT) and CPS connect with many smaller IoTs. They are far more complicated in design and applications than consumer-facing IoTs.

1.2.2 Big Data Analytics and Artificial Intelligence

Data analytics is one of the most important components in IoT's value chain. Big data in size and complexity, structured, semi-structured, and unstructured, outstrips the abilities to be processed by traditional data management systems.



Users

Consumers
Smart home system
Smart security
Smart entertainment
Smart healthcare…
Governments
Smart cities
Smart transportation
Smart energy
Smart grid…
Enterprises
Smart retail
Smart finance
Smart manufacturing
Smart agriculture…

5

FIGURE 1.1 Internet of Things ecosystem. Source: IDC, Amica Research

1.2.2.1 Big Data Characteristics

Big data has five main characteristics that are called five V's or volume, velocity, variety, veracity, and value.

Big data signifies a huge amount of data that is produced in a short period of time. A unit of measurement (UM) is entailed to define "big." The U.S. Library of Congress (LoC) is the largest library in the world that contains 167 million items occupying 838 miles (1,340 km) of bookshelves. This quantity of information is equivalent to 15 terabytes (TB) or 15×10^6 MB of digital data.[4] Using the contents of the Library of Congress as a UM is a good way to visualize the amount of information in 15TB of digital data.

Vast stream of data is being captured by AVs for navigation and analytics consequently to develop a safe and fully automated driving experience. AV collects data from cameras, lidars, sensors, and GPS that could exceed 4 TB of data per day. Tesla sold 368,000 AVs in 2019, which is 537,280,000 TB of data or 35,800 LoCs. This is only for one car model collected in 1 year. Considering data collected from all car models, airplanes, and devices in the universe, IDC forecasts there will be 163 ZB (1 ZB = 10^9 TB) of data by 2025, which is 10.9 million LoCs.

Velocity refers to speed at which new data is generated, analyzed, and moved around. Imagining AV navigation, social media message exchanges, credit card transaction execution, or high-frequency buying or selling stocks in milliseconds, the demands for execution must be immediate with high speed.

Variety denotes the different types of data. Structured data can be sorted and organized in tables or relational data-bases. The most common example is a table containing sales information by product, region, and duration. Nowadays the majority of data is unstructured data, such as social media conversations, photos, videos, voice recording, and sensor information that cannot fit into a table. Novel big data technology, including "Unstructured Data Management-as-a-Service," harnesses and sorts unstructured data into a structured manner that can be examined for relationships.

Veracity implies authenticity, credibility, and trustworthiness of the data. With big data received and processed at high speed, quality and accuracy of some data are at risk. They must be controlled to ensure reliable information is provided to users.

Last "v" but not least is value. Fast-moving big data in different variety and veracity is only useful if it has the ability to add value to users. It is imperative that big data analytics extracts business intelligence and adds value to data-driven management to make the right decision.

1.2.2.2 Data Analytics Anatomy

The IoT, mobile telecom, social media, etc. generate data with complexity through new forms, at high speed in real time and at a very large scale. Once the big data is sorted and organized using big data algorithms, the data are ready for analytical process (Fig. 1.2). The process starts from less sophisticated descriptive to highly sophisticated prescriptive analytics that ultimately brings value to users.

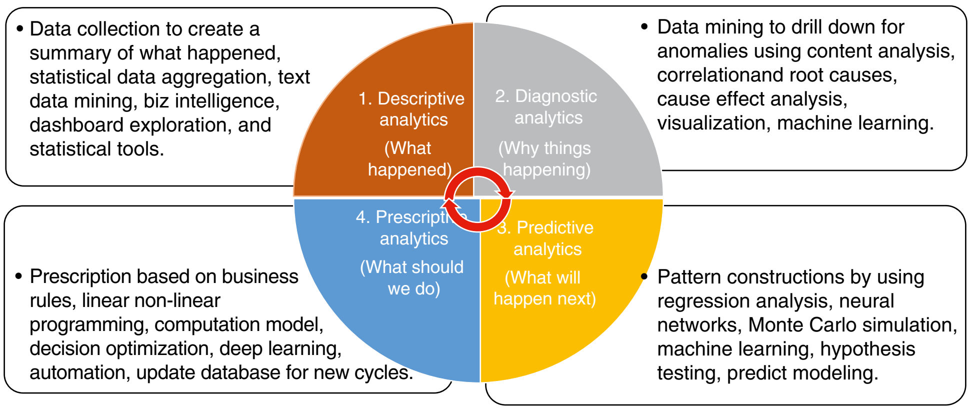


FIGURE 1.2 Virtuous Cycle of data analytics process with increasing difficulty and value. Source: © 2021 Amica Research.

[4] https://blogs.loc.gov/thesignal/2012/04/a-library-of-congress-worth-of-data-its-all-in-how-you-define-it/

Descriptive analytics does exactly what the name implies. It gathers historical data from relevant sources and cleans and transforms data into a proper format that a machine can read. Once the data is extracted, transformed, and loaded (ETL), data is summarized using data exploration, business intelligence, dashboard, and benchmark information.

Diagnostic analytics digs deeper into issues and finds in-depth root causes of a problem. It helps you understand why something happened in the past. Statistical techniques such as correlation and root cause, cause–effect analysis (Fig. 1.3), and graphic analytics visualize why the effect happened.

Predictive analytics helps businesses to forecast trends based on the current events. It predicts what is most likely to happen in the future and estimates time it will happen. Predictive analytics uses many techniques such as data mining, regression analysis, statistics, neural network, network analysis, predict modeling, Monte Carlo simulation, machine learning, etc.

Prescriptive analytics is the last and most sophisticated analytics that recommends what actions you can take to bring desired outcomes. It uses advanced tools such as decision tree, linear and nonlinear programming, deep learning, etc. to find optimal solutions and feedback to database for next analytics cycle.

Augmented analytics uses AI and machine learning to automate data preparation, discover insights, develop models and share insights among a broad range of business users. It is predicted that augmented analytics will be a dominant and destructive driver of data analytics.[5]

1.2.2.3 Artificial Intelligence

After years of fundamental research, AI is expanding and transforming every walk of life rapidly. AI has been used in IoT devices, autonomous driving, robot surgery, medical imaging and diagnosis, financial and economic modeling, weather forecasting, voice-activated digital assistance, and beyond. A well-designed AI application such as monitoring equipment failure and optimizing data center infrastructure operations and maintenance will save energy and avoid disasters.

John McCarthy, an assistant professor while at Dartmouth College, coined the term “artificial intelligence” in 1956. He defined AI as “getting a computer to do things which, when done by people, are said to involve intelligence.” There is no unified definition at the time of this publication, but AI technologies consist of hardware and software and the “machines that respond to simulation consistent with traditional responses from humans, given the human capacity of contemplation, judgment and intention.”[6] AI promises to drive from quality of life to the world economy. Applying both quantum computing, which stores information in 0’s, 1’s, or both called qubits, and parallel computing, which breaks a problem into discrete parts and solved many problems concurrently, AI can solve complicated problems faster and accurately in sophisticated ways and can conserve more energy in data centers.[7]

In data centers, AI could be used to monitor virtual machine operations and idle or running mode of servers, storages, and networking equipment to coordinate cooling loads and reduce power consumptions.

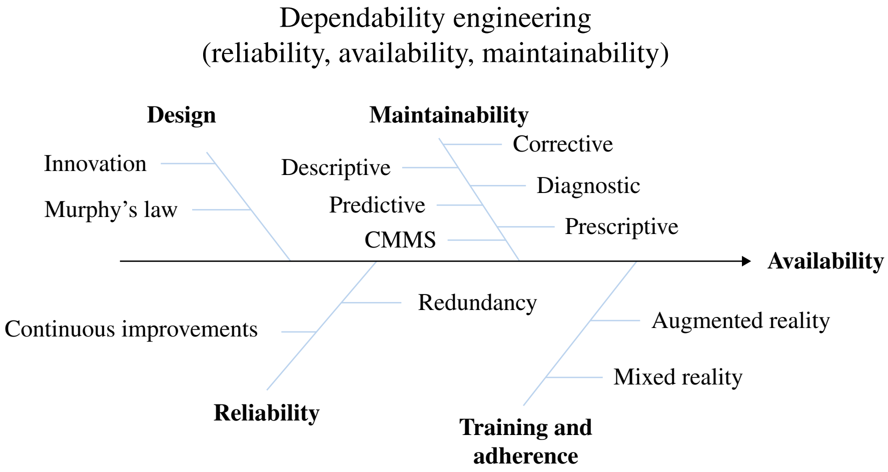


FIGURE 1.3 Cause and effect diagram. Source: © 2021 Amica Research.

[5] https://www.gartner.com/doc/reprints?id=1-1XOR8WDB&ct=191028&st=sb

[6] https://www.semanticscholar.org/paper/Applicability-of-Artificial-Intelligence-in-Fields-Shubhendu-Vijay/2480a71ef5e5a2b1f4a9217a0432c0c974c6c28c

[7] https://computing.llnl.gov/tutorials/parallel_comp/#Whatis

1.2.3 The Fifth-Generation Network

The 5G network, the fifth generation of wireless networks is changing the world and empowering how we live and work. 5G transmits median speed at 1.4 GB/s with reduced latency from 50 ms (1 ms = 0.001 s) to a few ms allowing little latency times for connected vehicles or remote surgery. There are wide spectra to provide 5G coverage. Using the high-frequency end of the spectrum, signals travel at extremely high speed, but the signals do not go as far nor through walls or obstruction. As a result, more wireless network equipment stations are required to be installed on streetlight or traffic poles. Using the lower-frequency end of the spectrum, signals travel farther but at a lower speed.

5G is one of the most important elements to power the IoT to drive smart manufacturing, smart transportation, smart healthcare, smart cities, smart entertainment, and smart everything. 5G can deliver incredibly detailed traffic, road, and hazard conditions to AV and power robotic surgery in real time. Through 5G, wearable glasses display patient's physical information and useful technical information to doctors in real time. 5G can send production instructions using wireless instead of wire at a faster speed that is critical to smart manufacturing. Virtual reality and augmented reality devices connect over 5G instead of wire that allows viewers to see the game from different angles in real time and superimpose player's statistics on the screen. By applying 5G, Airbus is piloting "Fello'fly or tandem flying" similar to migratory birds flying in a V formation to save energy.

1.3 DATA CENTER SYSTEM AND INFRASTRUCTURE ARCHITECTURE

The Oxford English dictionary defines architecture as "the art and study of designing buildings." The following are key components for architecture of a data center's system and infrastructure. They are discussed in detail in other chapters of this handbook.

- Mechanical system with sustainable cooling
- Electrical distribution and backup systems
- Rack and cabling systems
- Data center infrastructure management
- Disaster recovery and business continuity (DRBC)
- Software-defined data center
- Cloud and X-as-a-Service (X is a collective term referring to Platform, Infrastructure, AI, Software, DRBC, etc.)

1.4 STRATEGIC PLANNING

Strategic planning for data centers encompass a global location plan, site selection, design, construction, and operations. There is no one "correct way" to prepare a strategic plan. Depending on data center acquisition strategy (i.e., host, colocation, expand, lease, buy, build, or combination of above.), the level of deployments could vary from minor modifications of a server room to a complete build out of a green field project.

1.4.1 Strategic Planning Forces

The "Five Forces" described in Michael Porter's [3] "How Competitive Forces Shape Strategy" lead to a state of competition in all industries. The Five Forces are a threat of new entrants, bargaining power of customers, threat of substitute products or services, bargaining power of suppliers, and the industry jockeying for position among current competitors. Chinese strategist Sun Tzu, in the *Art of War*, stated five factors: the Moral Law, Weather, Terrain, the Commander, and Doctrine. Key ingredients in both strategic planning articulate the following:

- What are the goals
- What are fundamental factors
- What are knowns and unknowns
- What are constraints
- What are feasible solutions
- How the solutions are validated
- How to find an optimum solution

In preparing a strategic plan for a data center, Figure 1.4 shows four forces: business drivers, processes, technologies, and operations [4]. "Known" business drivers of a strategic plan include the following:

- Agility: Ability to move quickly.
- Resiliency: Ability to recover quickly from an equipment failure or natural disaster.
- Modularity and scalability: "Step and repeat" for fast and easy scaling of infrastructures.
- Reliability and availability: Ability of equipment to perform a given function and ability of an equipment to be in a state to perform a required function.
- Total cost of ownership (TCO): Total life cycle costs of CapEx (capital expenditures including land, building, design, construction, computer equipment, furniture and fixtures) and OpEx (operating expenditures including overhead, utility, maintenance, and repair costs).
- Sustainability: Apply best practices in green design, construction, and operations of data centers to reduce environmental impacts.

Additional "knowns" to each force could be expanded and added to tailor individual needs of a data center project. It is

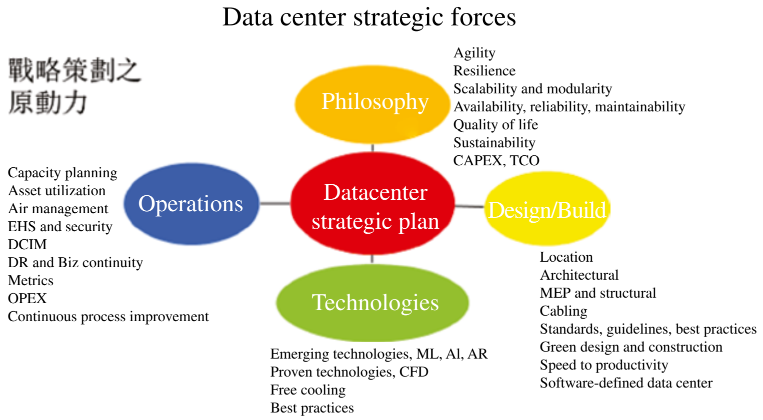


FIGURE 1.4 Data center strategic planning forces. Source: © 2021 Amica Research.

comprehensible that "known" business drivers are complicated and sometimes conflicting to each other. For example, increasing resiliency, or flexibility, of a data center will inevitably increase the costs of design and construction as well as continuous operating costs. The demand for sustainability will increase the TCO. "He can't eat his cake and have it too," so it is important to prioritize business drivers early on in the strategic planning process.

A strategic plan must anticipate the impacts of emerging technologies such as AI, blockchain, digital twin, and Generative Adversarial Networks, etc.

1.4.2 Capacity Planning

Gartner's study showed that data center facilities rarely meet the operational and capacity requirements of their initial design [5]. Microsoft's top 10 business practices estimated [6] that if a 12 Megawatt data center uses only 50% of power capacity, then every year $4–8 million in unused capital is stranded in uninterruptible power supply (UPS), generators, chillers, and other capital equipment. It is imperative to focus on capacity planning and resource utilization.

1.4.3 Strategic Location Plan

To establish data center location plan, business drivers include expanding market, emerging market, undersea fiber-optic cable, Internet exchange points, electrical power, capital investments, and many other factors. It is indispensable to have a strategic location roadmap on where to build data centers around the globe. Once the roadmap is established, a short-term data center design and implementation plan could follow. The strategic location plan starts from considering continents, countries, states, and cities down to a data center campus site. Considerations at continent and country or at macro level include:

- Political and economic stability of the country
- Impacts from political economic pacts (G20, G8+5, OPEC, APEC, RCEP, CPTPP, FTA, etc.)
- Gross domestic products or relevant indicators
- Productivity and competitiveness of the country
- Market demand and trend

Considerations at state (province) or at medium level include:

- Natural hazards (earthquake, tsunami, hurricane, tornado, volcano, etc.)
- Electricity sources with dual or multiple electrical grid services
- Electricity rate
- Fiber-optic infrastructure with multiple connectivity
- Public utilities (natural gas, water)
- Airport approaching corridor
- Labor markets (educated workforce, unemployment rate, etc.)

Considerations at city campus or at micro level include:

- Site size, shape, accessibility, expandability, zoning, and code controls
- Tax incentives from city and state
- Topography, water table, and 100-year floodplain
- Quality of life for employee retention
- Security and crime rate
- Proximity to airport and rail lines
- Proximity to chemical plant and refinery
- Proximity to electromagnetic field from high-voltage power lines
- Operational considerations

Other useful tools to formulate location plans include:

- Operations research
  - Network design and optimization
  - Regression analysis on market forecasting
- Lease vs. buy analysis or build and leaseback
- Net present value
- Break-even analysis
- Sensitivity analysis and decision tree

As a cross-check, compare your global location plan against data centers deployed by technology companies such as Amazon, Facebook, Google, Microsoft, and other international tech companies.

1.5 DESIGN AND CONSTRUCTION CONSIDERATIONS

A data center design encompasses architectural (rack layout), structural, mechanical, electrical, fire protection, and cabling system. Sustainable design is essential because a data center can consume 40–100 times more electricity compared to a similar-size office space. In this section, applicable design guidelines and considerations are discussed.

1.5.1 Design Guidelines

Since a data center involves 82–85% of initial capital investment in mechanical and electrical equipment [7], data center project is generally considered as an engineer-led project. Areas to consider for sustainable design include site selection, architectural/engineering design, energy efficiency best practices, redundancy, phased deployment, etc. There are many best practices covering site selection and building design in the Leadership in Energy and Environmental Design (LEED) program. The LEED program is a voluntary certification program that was developed by the U.S. Green Building Council (USGBC).[8]

Early on in the architecture design process, properly designed column spacing and floor elevation will ensure appropriate capital investments and minimize operating expenses. A floor plan with appropriate column spacing maximizes ICT rack installations and achieves power density with efficient cooling distribution. A floor-to-floor elevation must be carefully planned to include height and space for mechanical, electrical, structural, lighting, fire protection, and cabling system.

International technical societies have developed many useful design guidelines that are addressed in detail in other chapters of this handbook:

- ASHRAE TC9.9: Data Center Networking Equipment [8]
- ASHRAE TC9.9: Data Center Power Equipment Thermal Guidelines and Best Practice
- ASHRAE 90.1: Energy Standard for Buildings [9]
- ASHRAE: Gaseous and Particulate Contamination Guidelines for Data Centers [10]
- Best Practices Guide for Energy-Efficient Data Center Design [11]
- EU Code of Conduct on Data Centre Energy Efficiency [12]
- BICSI 002: Data Center Design and Implementation Best Practices [13]
- FEMA P-414: "Installing Seismic Restraints for Duct and Pipe" [14]
- FEMA 413: "Installing Seismic Restraints for Electrical Equipment" [15]
- FEMA, SCE, VISCMA, "Installing Seismic Restraints for Mechanical Equipment" [16]
- GB 50174: Code for Design of Data Centers [17]
- ISO 50001: Energy Management Specification and Certification
- LEED Rating Systems [18]
- Outline of Data Center Facility Standard by Japan Data Center Council (JDCC) [19]
- TIA-942: Telecommunications Infrastructure Standard for Data Centers

Chinese standard GB 50174 "Code for Design of Data Centers" provides a holistic approach of designing data centers that cover site selection and equipment layout, environmental requirements, building and structure, air conditioning (mechanical system), electrical system, electromagnetic shielding, network and cabling system, intelligent system, water supply and drainage, and fire protection and safety [17].

1.5.2 Reliability and Redundancy

"Redundancy" ensures higher reliability, but it has profound impacts on initial investments and ongoing operating costs (Fig. 1.3).

In 2011, with fierce competition against Airbus SE, Boeing Company opted to update its single-aisle 737 rather than design a new jet that is equipped with new fuel-efficient engines. The larger engines were placed farther forward on the wing that, in certain condition, caused the plane nose to pitch up too quickly. The solution to the problem was to use MCAS (Maneuvering Characteristics Augmentation System) that is a stall prevention system. For the 737 Max, a single set of "angle-of-attack" sensors was used to determine if automatic flight control commands should be triggered when the MCAS is fed sensor data. If a second set of sensors

[8] http://www.usgbc.org/leed/rating-systems

and software or redundancy design on angle of attack had been put in place, two plane crashes, which killed 346 people 5 months apart, could have been avoided [20, 21].

Uptime Institute® pioneered a tier certification program that structured data center redundancy and fault tolerance in four tiers [22]. Telecommunication Industry Association's TIA-942 contains four tables that describe building and infrastructure redundancy in four levels. Basically, different redundancies are defined as follows:

- N: Base requirement.
- $N+1$ redundancy: Provides one additional unit, module, path, or system to the minimum requirement
- $N+2$ redundancy: Provides two additional units, modules, paths, or systems in addition to the minimum requirement
- $2N$ redundancy: Provides two complete units, modules, paths, or systems for every one required for a base system
- $2(N+1)$ redundancy: Provides two complete $(N+1)$ units, modules, paths, or systems

Accordingly, a matrix table is established using the following tier levels in relation to component redundancy:

Tier I Data Center: Basic system
Tier II Data Center: Redundant components
Tier III Data Center: Concurrently maintainable
Tier V Data Center: Fault tolerant

The China National Standard GB 50174 "Code for Design of Data Centers" defines A, B, and C tier levels with A being the most stringent.

JDCC's "Outline of Data Center Facility Standard" tabulates "Building, Security, Electric Equipment, Air Condition Equipment, Communication Equipment and Equipment Management" in relation to redundancy Tiers 1, 2, 3, and 4. It is worthwhile to note that the table also includes seismic design considerations with probable maximum loss (PML) relating to design redundancy.

Data center owners should consult and establish a balance between desired reliability, redundancy, PML, and additional costs.[9]

1.5.3 Computational Fluid Dynamics

Whereas data centers could be designed by applying best practices, the locations of systems (rack, CRAC, etc.) might not be in its optimal arrangement collectively. Computational fluid dynamics (CFD) technology has been used in semiconductor's cleanroom projects for decades to ensure uniform airflow inside a cleanroom. During the initial building and rack layout design stage, CFD offers a scientific analysis and solution to visualize airflow patterns and hot spots and validate cooling capacity, rack layout, and location of cooling units. One can visualize airflow in hot and cold aisles for optimizing room design. During the operating stage, CFD could be used to emulate and manage airflow to ensure the air path does not recirculate, bypass, or create negative pressure flow.

1.5.4 Best Practices

Although designing energy-efficient data center is still evolving, many best practices could be applied whether you are designing a small server room or a large data center. One of the best practices is to build or use ENERGY STAR servers [23] and solid-state drives. The European Commission published a comprehensive "Best Practices for the EU Code of Conduct on Data Centres." The U.S. Department of Energy's Federal Energy Management Program published "Best Practices Guide for Energy-Efficient Data Center Design." Both, and many other publications, could be referred to when preparing a data center design specification. Here is a short list of best practices and emerging technologies:

- In-rack-level liquid cooling and liquid immersion cooling
- Increase server inlet temperature and humidity adjustments (ASHRAE Spec) [24]
- Hot and cold aisle configuration and containment
- Air management (to stop bypass, hot and cold air mixing, and recirculation)
- Free cooling using air-side economizer or water-side economizer
- High efficient UPS
- Variable speed drives
- Rack-level direct liquid cooling
- Fuel cell technology
- Combined heat and power (CHP) in data centers [22]
- Direct current power distribution
- AI and data analytics applications in operations control.

It is worthwhile to note that servers can operate outside the humidity and temperature ranges recommended by ASHRAE [25].

1.6 OPERATIONS TECHNOLOGY AND MANAGEMENT

Best practices in operations technology (OT) and management include benchmark metrics, data center infrastructure management, air management, cable management,

[9] www.AmicaResearch.org